* What are the feature vectors?

A feature vector is an n-dimensional vector of numerical features that represent an object. In machine learning, feature vectors are used to represent numeric or symbolic characterictics(called features) of an object in a mathematical way that’s easy to analyze.

* What is root cause analysis?

Root cause analysis was initially developed to analyze industrial accidents but is now widely used in other areas. It is a problem-solving technique used for isolating the root causes of faults or problems. A factor is called a root cause if its deduction from the problem-faults-sequence averts the final undesirable event from recurring.

* What is logistic regression?

Logistic regression is also known as the logit model. It is a technique used to forecast the binary outcome from a linear combination of predictor variables.

* What are recommender systems?

Recommender systems are a subclass of information filtering systems that are meant to predict the preferences or ratings that a user would give to a product.

* Explain Cross-validation.

Cross-validation is a model validation technique for evaluating how the outcomes of a statistical analysis will generalize to an independent data set. It is mainly used in backgrounds where the objective is to forecast and one wants to estimate how accurately a model accomplish in practice.

The goal of cross-validation is to term a data set to test the model in the training phase (i.e. validation data set) to limit problems like overfitting and gain insight into how the model will generalize to an independent data set.

* What is collaborative filtering?

Most recommender systems use this filtering process to find patterns and information by collaborating perspectives, numerous data sources, and several agents.

* Do gradient descent methods always converge to similar points?

They do not, because in some cases, they reach a local minima or a local optima point. You would not reach the global optima point. This is governed by the data and the starting conditions.

* What is the goal of A/B Testing?

This is statistical hypothesis testing for randomized experiments with two variables, A and B. The objective of A/B testing is to detect any changes to a web page to maximize or increase the outcome of strategy.

* What are the drawbacks of the linear model?

1. The assumption of linearity of the errors.
2. It can’t be used for count outcomes or binary outcomes.
3. There are overfitting problems that it can’t solve

* What is law of large numbers?

It is a theorem that describes the result of performing the same experiment very frequently. This theorem forms the basis of frequency-style thinking. It states that the sample mean, sample variance , and sample standard deviation converge to what they are trying to estimate.

* What are the confounding variables?

These are extraneous variables in a statistical model that correlates directly or inversely with bith the dependent and the independent variable. The estimate fails to account for the confounding factor.

* What is star schema?

It is a traditional database schema with a central table. Satellite tables map IDs to physical names or descriptions and can be connected to the central fact table using the IDs fields; these tables are knows as lookup tables and are principally useful in real-time applications, as they save a lot of memory. Sometimes, star schemas involve several layers of summarization to recover information faster.

* What are eigenvalue and eigenvector ?

Eignevalues are the directions along which a particular linear transformations acts by flipping, compressing or stretching.

Eigenvectors are for understanding linear transformations. In data analysis, we usually calculate the eigenvectors for a correlation or covariance matrix.

* Why is resampling done?

Resampling is done in any of these cases:

1. Estimating the accuracy of sample statistics by using subsets of accessible data, or drawing randomly with replacement from a set of data points.
2. Substituting labels on data points when performing significance tests.
3. Validating models by using random subsets (bootstrapping, cross-validation).